5-2009

Using high-frequency transaction data to estimate the probability of informed trading

Anthony S. TAY
Singapore Management University, anthonytay@smu.edu.sg

Christopher TING
Singapore Management University, christophert@smu.edu.sg

Yiu Kuen TSE
Singapore Management University, yktse@smu.edu.sg

Mitchell WARACHKA
Singapore Management University, mitchell@smu.edu.sg

DOI: https://doi.org/10.1093/jjfinec/nbp005

Published in Journal of Financial Econometrics (2009) 7 (3): 288-311.
http://doi.org/10.1093/jjfinec/nbp005

Using High-Frequency Transaction Data to Estimate the Probability of Informed Trading

ANTHONY TAY
Singapore Management University

CHRISTOPHER TING
Singapore Management University

YIU KUEN TSE
Singapore Management University

MITCH WARACHKA
Singapore Management University

ABSTRACT
This paper applies the asymmetric autoregressive conditional duration (AACD) model of Bauwens and Giot (2003) to estimate the probability of informed trading (PIN) using irregularly spaced transaction data. We model trade direction (buy versus sell orders) and the duration between trades jointly. Unlike the Easley, Hvidkjaer, and O'Hara (2002) approach, which uses the aggregate numbers of daily buy and sell orders to estimate PIN, our methodology allows for interactions between consecutive buy-sell orders and accounts for the duration between trades and the volume of trade. We extend the Easley–Hvidkjaer–O'Hara framework by allowing the probabilities of good news and bad news to vary each day. Our PIN estimates can be computed daily as well as over intraday intervals. (*JEL*: C410, G120)

KEYWORDS: autoregressive conditional duration, market microstructure, probability of informed trading, transaction data, Weibull distribution

The authors are indebted to the associate editor and two anonymous referees for their many helpful comments and suggestions. Any remaining errors are the responsibility of the authors. Tao Yang provided excellent research assistance. Yiu Kuen Tse gratefully acknowledges research support from MOE Tier 2 research grant T206B4301-RS. Address correspondence to Yiu Kuen Tse, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, or e-mail: yktse@smu.edu.sg.

This paper proposes a method to estimate the probability of informed trading (PIN) using high-frequency transaction data. Our method is based on the asymmetric autoregressive conditional duration (AACD) model of Bauwens and Giot (2003). Modeling trade direction (buyer- versus seller-initiated trades) and the duration of these transactions *jointly*, we estimate the intensity of informed versus uninformed trading. These estimates are then used to compute PIN. In contrast to the Easley, Hvidkjaer, and O'Hara (2002) (EHO hereafter) framework, which uses the aggregate numbers of daily buy and sell orders to estimate PIN, our methodology uses irregularly spaced transaction data.

In the EHO framework, PIN is estimated using daily aggregates of buy and sell orders, which are assumed to be independent. The probabilities of no news, good news, and bad news are assumed to be constant, and volume is not taken into account. We relax these assumptions and use high-frequency transaction data on buy and sell orders to model trade directions and trade durations using the AACD framework.

Easley, Kiefer, and O'Hara (1996, 1997) established the theoretical foundations of PIN. Since then, many empirical applications of PIN have appeared in the finance literature. For example, Easley, Kiefer, and O'Hara (1996) used PIN to investigate the role of purchased order flow. Easley et al. (1996) and Easley, O'Hara, and Saar (2001) employ PIN to study infrequently traded stocks and the impact of stock splits, respectively. Aslan et al. (2006) use PIN to examine the link between market microstructure and asset pricing. Henry (2006) estimates PIN to study the interaction between short selling and informed trading. Chung and Li (2003) utilize PIN to verify the appropriateness of decomposing bid-ask spreads into adverse-selection and transitory components, while Lei and Wu (2005) find PIN predicts various bid-ask measures. Recently, Duarte and Young (2009) demonstrate that PIN can be decomposed into two components, which are related to asymmetric information and illiquidity. They examine the relation between PIN and the cross-section of expected returns.

The widespread applications of PIN have recently led many researchers to examine critically its properties. Aktas et al. (2007) find that the behavior of PIN contradicts the leakage of information around merger-acquisition announcements. Benos and Jochec (2007) report that PIN, estimated under the EHO method, is lower before earnings announcements. They argue that this "anomalous" behavior is due to the assumption of constant probabilities of news in the EHO model, as well as its failure to account for transaction volume.

Engle and Russell (1998) and Engle (2000) propose the autoregressive conditional duration (ACD) model to analyze the duration between two transactions, irrespective of whether they correspond to a price increase or decrease and whether they are initiated by a buy or sell order. Bauwens and Giot (2003) extend the ACD model to study the mid-price of bid-ask quotes. They propose a two-state AACD model to analyze mid-price decreases and increases jointly with trade duration. In their model, the conditional expected duration of each state varies with conditional information. This conditional information can include lagged durations, lagged volume, and the lagged spread. Recently, the ACD literature has rapidly expanded, with contributions by Bauwens and Veredas (2004), Fernandes and Grammig (2005), Ghysels, Gourieroux, and Jasiak (2004), Grammig and Maurer (2000), Bisière and Kamionka (2000), and Zhang, Russell, and Tsay (2001), among others. Pacurar (2008) provides a comprehensive survey of ACD models.[1]

We apply the AACD approach to a two-state model of transaction data, where the two states represent a transaction initiated by either a buy or a sell order. Following Bauwens and Giot (2003), we allow the expected duration to vary with covariates that include lagged duration, lagged conditional expected duration, lagged trade direction, and lagged trade volume. We construct AACD equations that reflect changes in trade intensity. Using transaction data, we are able to model the interaction between buy and sell orders, and estimate PIN on a daily basis as well as over intraday intervals. Furthermore, we propose a model in which the probabilities of no news, good news, and bad news vary each day. The resulting estimates of PIN avoid many of the restrictions in EHO.

The remainder of the paper is as follows. In Section 1, we review briefly the EHO framework for PIN and their estimation method. In Section 2 we outline our proposed AACD model of trade direction and duration, assuming constant probabilities of news. PIN is estimated as the ratio of the aggregate intensity of informed trades to the total intensity of all trades, weighted by their transaction durations. We also incorporate time-varying probabilities for good news and bad news. Sections 3 and 4 describe the data used in our empirical illustration and empirical results, respectively. Our conclusions and suggestions for future research are contained in Section 5.

1 PROBABILITY OF INFORMED TRADING

EHO use trade-direction data to estimate the proportion of trades initiated by informed traders. Their model is based on the number of buy and sell orders, the intensities of which depend on the existence of "news" or information. Conditional on the arrival of news, information is further classified as being either "good" or "bad". EHO model the aggregate number of buy- and sell-initiated trades each day as independent Poisson variables, with different intensities for days with good news, no news and bad news. The characterization of each trading day is unknown, and the likelihood of the numbers of buy and sell orders is based on the mixture-of-Poisson distribution. PIN is then calculated as the ratio for the intensity of informed trades divided by the intensity of all trades (informed or uninformed).

In the EHO framework, each trading day is characterized by good news (G), no news (N), and bad news (B) to form the set $S = \{G, N, B\}$. We denote π_s as the probability of state s in S. Let the probability of a day containing news be θ_E. Conditional on the arrival of news, we denote the probability of bad news by θ_B.

[1] An alternative approach that parallels the ACD model is the autoregressive conditional intensity (ACI) model. See, e.g., Russell (1999), Bauwens and Hautsch (2006), and Hautsch (2004).

Thus, the probability of a no-news day is $\pi_N = 1 - \theta_E$, and the probabilities of good- and bad-news days are $\pi_G = \theta_E\,(1-\theta_B)$ and $\pi_B = \theta_E\,\theta_B$, respectively. EHO assume the aggregate numbers of buy and sell orders on a trading day follow independent Poisson distributions, where the intensities of sell and buy orders on a no-news day, denoted by λ_{-1} and λ_1, respectively, are constant throughout the sample period. On a good-news day, the buy intensity increases by a positive amount δ, with no change in the sell intensity. Likewise, on a bad-news day, the sell intensity increases by δ while the buy intensity remains unchanged. With D days of data, the mixture-of-distributions assumption implies that the likelihood function equals

$$\prod_{d=1}^{D}\left[(1-\theta_E)\,\frac{\lambda_1^{B_d}e^{-\lambda_1}}{B_d!}\,\frac{\lambda_{-1}^{S_d}e^{-\lambda_{-1}}}{S_d!}+\theta_E\,\theta_B\,\frac{\lambda_1^{B_d}e^{-\lambda_1}}{B_d!}\,\frac{(\lambda_{-1}+\delta)^{S_d}\,e^{-(\lambda_{-1}+\delta)}}{S_d!}\right.$$
$$\left.+\,\theta_E\,(1-\theta_B)\frac{(\lambda_1+\delta)^{B_d}\,e^{-(\lambda_1+\delta)}}{B_d!}\,\frac{\lambda_{-1}^{S_d}\,e^{-\lambda_{-1}}}{S_d!}\right], \tag{1}$$

where B_d and S_d are the respective aggregate number of buy and sell orders on day d. From this model, EHO estimate PIN as

$$\text{PIN} = \frac{\theta_E\,\delta}{\theta_E\,\delta + \lambda_{-1} + \lambda_1}. \tag{2}$$

In other words,

$$\text{PIN} = \frac{\text{Expected number of trades per day initiated by informed traders}}{\text{Expected total number of trades per day}}. \tag{3}$$

Equation (3) may be adopted as the enhanced definition of PIN. First, it defines PIN directly by the *activity* (number of trades) of the informed and uninformed traders. Second, it reduces to the original PIN defined by EHO in Equation (2) if the Poisson assumption of trade frequency is adopted. Third, it allows researchers to use alternative assumptions for the arrival of trades. For example, inter-arrival time of trades may be assumed to be Weibull (instead of exponential as in the EHO case), and Equation (3) enables us to compute the PIN under such alternative assumptions.

The literature that adopts the EHO methodology typically computes PIN over monthly or annual intervals, during which PIN and the arrival of news are assumed to be constant. Thus, any possible variation in PIN is ignored. According to Aktas et al. (2007), at least two months of daily observations are required for the maximum likelihood estimation in EHO to converge. This appears to subject the EHO methodology to some serious restrictions. Furthermore, these authors highlight the importance of volume to the estimation of PIN. They argue that ignoring volume causes PIN to be insensitive to market trends.

Recently, some studies estimated PIN at higher frequencies, such as on a daily basis. Lei and Wu (2005) study the explanatory power of PIN on various measures of stock spreads in a panel regression of daily data. They estimate PIN over quarterly periods and assign the same PIN to each trading day within this horizon. Their study highlights the importance of high-frequency PIN, although the estimation method appears to be *ad hoc*. Easley et al. (2008) propose a bivariate time series approach for the estimation of PIN. They model the arrival of informed and uninformed trades using a bivariate autoregressive process, with total trade and trade imbalance (difference between buy and sell orders) as covariates. To identify the parameters in their model, they assume that the probability of an information event is constant, while the informed and uninformed trades are time varying.[2]

We attempt to extend the PIN literature by proposing a methodology to estimate PIN using transaction data. The use of transaction data enables us to compute PIN on a daily basis as well as over intraday intervals. Also, we are able to relax the assumption of constant probabilities of news and avoid imposing this restriction over an extended period of time. These flexibilities are not found in the Easley–Engle–O'Hara–Wu approach.

2 HIGH-FREQUENCY PIN ESTIMATION

This section proposes a methodology to estimate PIN using high-frequency transaction data. Our method is based on an AACD model for trade directions. We first outline the AACD model as proposed by Bauwens and Giot (2003). The specification of the conditional expected duration equation is then constructed to account for informed and uninformed trades due to the arrival of news. Finally, we extend our methodology to allow the probabilities of good and bad news to vary across time, and estimate PIN on a daily basis.

2.1 The AACD Model of Trade Direction

We model trade directions using a two-state AACD model, and compute PIN from this model. Previous AACD model applications such as Bauwens and Giot (2003) involve prices rather than trade directions. Lo and Sapp (2005) apply the AACD model to study the interaction between limit and market orders.

Let y_i denote the trade direction of the ith trade at time t_i, which may take values of $j = -1$ or 1 representing a sell-initiated and buy-initiated trade, respectively. We denote Φ_{i-1} as the information set after the $(i-1)$th trade. Φ_{i-1} may consist of past trade directions, transaction volume, and lagged durations. Given Φ_{i-1}, we assume each of the two *potential* trade directions of the trade at time t_i follows a latent stochastic point process whose inter-arrival times have independent exponential distributions. The realized (observed) trade direction is the outcome of the competition between the two underlying Poisson point processes to be the first arrival.

Specifically, conditional on the information set Φ_{i-1} after the $(i-1)$th trade, the inter-arrival time random variables of the latent processes follows an exponential

[2]Given that Easley et al. (2008) use more than 15 years of daily data, it is doubtful whether this assumption can be justified.

distribution with mean (conditional expected duration) ψ_{ji}, where j denotes the latent trade direction and i denotes the trade at time t_i. We denote $\lambda_{ji} = 1/\psi_{ji}$, which is the *intensity* of the latent Poisson process. We further denote the duration of the ith trade (i.e., the waiting time from time t_{i-1} of the $(i-1)$th trade to time t_i of the ith trade) by $x_i = t_i - t_{i-1}$, the conditional joint density of (x_i, y_i) (or equivalently (t_i, y_i)), denoted by $p_i(x_i, y_i|\Phi_{i-1})$, is then given by[3]

$$p_i(x_i, j|\Phi_{i-1}) = \lambda_{ji} \exp\left[-(\lambda_{-1,i} + \lambda_{1i})\, x_i\right], \qquad j = -1, 1. \tag{4}$$

Summing over the possible trade directions j in Equation (4), we obtain the following conditional marginal density of x_i:

$$f_{x_i}(x|\Phi_{i-1}) = (\lambda_{-1,i} + \lambda_{1i})\, \exp\{-(\lambda_{-1,i} + \lambda_{1i})x\}, \tag{5}$$

which is an exponential distribution with mean $1/(\lambda_{-1,i} + \lambda_{1i})$. On the other hand, integrating $p_i(x_i, y_i|\Phi_{i-1})$ over the duration x_i, the following conditional marginal density of y_i is obtained:

$$\begin{aligned} f_{y_i}(j|\Phi_{i-1}) &= \int_0^\infty \lambda_{ji} \exp\{-(\lambda_{-1,i} + \lambda_{1i})x\}\, dx \\ &= \frac{\lambda_{ji}}{\lambda_{-1,i} + \lambda_{1i}}, \qquad j = -1, 1\,. \end{aligned} \tag{6}$$

As the joint density in Equation (4) is the product of the marginal density of x_i and y_i, x_i and y_i are independent conditional on the information set Φ_{i-1}.

Given a sample of observations $\{x_i, y_i\}$ for $i = 1, \ldots, N$, the log-likelihood function may be written as

$$\sum_{i=1}^{N} \log\, p_i(x_i, y_i|\Phi_{i-1}) = -\sum_{i=1}^{N} \left[\sum_{j=-1,1} \frac{x_i}{\psi_{ji}} - \log\left(\sum_{j=-1,1} \frac{D_{y_i}(j)}{\psi_{ji}} \right) \right], \tag{7}$$

where $D_{y_i}(j) = 1$, if $j = y_i$ and 0 otherwise. Thus, the parameters of the model can be estimated using maximum likelihood estimation (MLE) once the functional forms of the conditional expected durations ψ_{ji} are specified. For this purpose, we adopt the ACD model of Engle and Russell (1998). For example, the conditional expected duration may be specified in the following logarithmic form (see Bauwens and Giot 2000),

$$\log \psi_{ji} = \nu_{j,-1} D_{-1}(y_{i-1}) + \nu_{j1} D_1(y_{i-1}) + \alpha_j \log \psi_{j,i-1} + \beta_j \log x_{i-1}, \qquad j = -1, 1. \tag{8}$$

In the above equation, we have an extended logarithmic ACD(1, 1) structure, where the constant term in the usual ACD equation is replaced by the intercepts

[3]Refer to Equation (11) of Bauwens and Giot (2003), which is given in slightly different notations.

$\nu_{j,-1}$ and ν_{j1} that vary according to the previous trade direction y_{i-1}. An increase (decrease) in ψ_{ji} implies a larger (smaller) expected duration, which in turn implies a reduced (increased) probability that the transaction at time t_i is of type j. The intercepts ν_{jk} represent the sensitivity of the next trade direction j to the prior trade direction k. Thus, if the previous trade direction is of type k, the intercept for $\log \psi_{ji}$ is ν_{jk}. A larger (smaller) ν_{jk} implies that trade direction k induces a larger (smaller) expected duration of the next trade direction being of type j.

We conclude this section by mentioning some possible extensions of the above model. First, as pointed out by Aktas et al. (2007), the EHO estimate of PIN does not incorporate transaction volume. We shall propose ACD equations that allow trade volume to affect the conditional expected duration, hence the trading intensity. Second, the exponential assumption for inter-arrival time may be relaxed to include more general distributional assumptions, such as the two-parameter Weibull distribution. The Appendix provides an outline of the key results for the Weibull case. However, the estimation results of the AACD model with the Weibull assumption are found to be similar to those under the exponential assumption. Thus, we maintain the exponential assumption as it provides PIN formulas that are easier to interpret.

2.2 Constant Probabilities of News

We now insert the AACD model into the EHO framework by specifying the conditional expected duration equation as being dependent on the state of good news, no news, or bad news. First, we denote ψ^s_{ji} as the conditional expected duration of trade direction j in state $s \in S$ given information Φ_{i-1}, where the specification of ψ^s_{ji} reflects the activities of informed and uninformed traders. We define the following function f^s_{ji} as the basis of the equations for the conditional expected duration in each of the three states in S:

$$f^s_{ji} \equiv \nu_{j,-1} D_{-1}(y_{i-1}) + \nu_{j1} D_1(y_{i-1}) + \alpha_j \ \log \psi^s_{j,i-1} + \beta_j \ \log x_{i-1} + \varsigma_j \ y_{i-1} \log v_{i-1}\,, \tag{9}$$

for $j = -1, 1$ and $s \in S$, where v_{i-1} is the volume of the trade at time t_{i-1}. Thus, the basis f^s_{ji} depends on whether the previous transaction is a buy- or sell-initiated order y_{i-1}, the lagged duration x_{i-1}, the previous conditional expected duration $\psi^s_{j,i-1}$, as well as the lagged signed logarithmic volume $y_{i-1} \log v_{i-1}$. Hence, Equation (9) allows volume to impact trade intensity.

According to the assumptions of EHO, only uninformed traders are active in the absence of any news. When there is good news, informed traders purchase shares, increasing the trading intensity of buy orders. Conversely, when there is bad news, informed traders sell shares and increase the trading intensity of sell orders. However, the trading intensity of sell orders on a good-news day and the trading intensity of buy orders on a bad-news day are identical to their counterparts on a no-news day.

Thus, for a no-news day ($s = N$), we assume that the basic functional form for the logarithmic conditional expected duration in Equation (9) holds, so that

$$\log \psi_{ji}^{N} = f_{ji}^{N}, \qquad j = -1, 1. \tag{10}$$

For the buy-orders ($j = 1$) on a good-news day ($s = G$), we reduce f_{1i}^{N} by a positive constant μ to yield the following logarithmic conditional expected duration:

$$\log \psi_{1i}^{G} = f_{1i}^{G} - \mu, \tag{11}$$

while the logarithmic conditional expected duration for a sell trade is the basis function $f_{-1,i}^{G}$,

$$\log \psi_{-1,i}^{G} = f_{-1,i}^{G}. \tag{12}$$

Conversely, on a bad-news day ($s = B$), we have

$$\log \psi_{1i}^{B} = f_{1i}^{B} \tag{13}$$

and

$$\log \psi_{-1,i}^{B} = f_{-1,i}^{B} - \mu. \tag{14}$$

According to Equation (13), the logarithmic conditional expected duration of a buy order on a bad-news day is the benchmark function f_{1i}^{B}. However, the logarithmic conditional expected duration of a sell order $\log \psi_{-1,i}^{B}$ on a bad-news day decreases by μ due to selling by informed traders. As seen in Equation (11), on a good-news day, the logarithmic conditional expected duration of a buy order decreases due to buying by informed traders. However, the logarithmic conditional expected duration of sell orders on good-news days, $\log \psi_{-1,i}^{G}$, remain unchanged versus that of a no-news day, as seen in Equation (12).[4]

Given that a certain day is of type s, the joint density of (x_i, y_i) conditional on the information set Φ_{i-1} is given in Equation (4), which is rewritten below to incorporate variation with respect to the state of news:

$$p_{si}(x_i, k|\Phi_{i-1}) = \prod_{j=-1,1} \left(\frac{1}{\psi_{ji}^{s}} \right)^{D_k(j)} \exp\left(-\frac{x_i}{\psi_{ji}^{s}} \right), \qquad k = -1, 1;\ s \in S. \tag{15}$$

Let $N_d = S_d + B_d$ denote the number of trades on day d. The likelihood function is then given by

$$\prod_{d=1}^{D} \left[\sum_{s \in S} \pi_s \left(\prod_{i=1}^{N_d} p_{si}(x_i, y_i|\Phi_{i-1}) \right) \right]. \tag{16}$$

[4] Note that we may further enhance the flexibility of the model by allowing the reduction in the expected duration μ to differ in buy versus sell orders, as might be justified due to short-selling restrictions (see Diamond and Verrecchia 1991). We are indebted to an anonymous referee for suggesting this possibility.

Note that the product term in the inner brackets of Equation (16) is the likelihood function for day d, given that day d is in state s (the index d for the $\{x_i, y_i\}$ data is suppressed).

In the PIN–EHO model, the Poisson assumption is adopted so that the hazard rate is constant and is equal to the reciprocal of the expected duration, which is used as a measure of the *intensity*. Thus, EHO define PIN as the ratio of the expected number of trades originated by informed traders divided by the total expected number of trades, whether they are initiated by informed or uninformed traders.

In the AACD model, the expected number of trades due to informed traders in the interval (t_{i-1}, t_i), conditional on Φ_{i-1}, equals

$$\left(\pi_G\, \lambda_{1i}^G + \pi_B\, \lambda_{-1,i}^B\right) x_i. \tag{17}$$

Likewise, the expected number of trades due to all traders in the interval (t_{i-1}, t_i), conditional on Φ_{i-1}, is

$$\left(\lambda_{-1,i}^N + \lambda_{1i}^N + \pi_G\, \lambda_{1i}^G + \pi_B\, \lambda_{-1,i}^B\right) x_i. \tag{18}$$

As shown in Equation (3), PIN in the EHO framework is the proportion of the expected number of trades due to informed traders to the total expected number of all trades. Thus, aggregating the quantities in Equations (17) and (18) over the entire period of the sample, we compute PIN as the ratio of the total expected number of trades due to informed traders to the total expected number of all trades over all trading intervals, i.e.,

$$\text{PIN} = \frac{\sum_{d=1}^{D} \sum_{i=1}^{N_d} \left(\pi_G\, \lambda_{1i}^G + \pi_B\, \lambda_{-1,i}^B\right) x_i}{\sum_{d=1}^{D} \sum_{i=1}^{N_d} \left(\lambda_{-1,i}^N + \lambda_{1i}^N + \pi_G\, \lambda_{1i}^G + \pi_B\, \lambda_{-1,i}^B\right) x_i}, \tag{19}$$

where again the index d for the intensities and the $\{x_i, y_i\}$ data is suppressed.

Equation (19) is a generalization of Equation (2) with time-varying trade intensities. In the special case where the intensities are constant, whether there is news or no news, we have: $\lambda_{1i}^G = \lambda_{-1,i}^B = \delta$ (assuming the trade intensities of informed traders under good news and bad news are the same, as in the EHO framework), $\lambda_{1i}^N = \lambda_1$, $\lambda_{-1,i}^N = \lambda_{-1}$ and $\pi_G + \pi_B = \theta_E$. From this, it can be easily seen that Equation (19) reduces to Equation (2).

Although Equation (19) offers a PIN estimate over the entire sample period, the formula can be modified to estimate PIN on a specific day. Denoting PIN_d as PIN on day d, we can estimate PIN_d by

$$\text{PIN}_d = \frac{\sum_{i=1}^{N_d} \left(\pi_G\, \lambda_{1i}^G + \pi_B\, \lambda_{-1,i}^B\right) x_i}{\sum_{i=1}^{N_d} \left(\lambda_{-1,i}^N + \lambda_{1i}^N + \pi_G\, \lambda_{1i}^G + \pi_B\, \lambda_{-1,i}^B\right) x_i}, \tag{20}$$

where the data $\{x_i, y_i\}$ and the estimated parameters $\lambda_{-1,i}^N$, λ_{1i}^N, λ_{1i}^G, and $\lambda_{-1,i}^B$ pertain to day d. Indeed, Equation (20) can be used to compute PIN measures over intraday intervals, in which case the summations in the numerator and denominator are over trades in subintervals of a day.

In summary, we have introduced an AACD model of trade direction, which maintains the EHO assumption that the probabilities of good news, no news, and bad news are constant in the sample period. Using this model, the *average* PIN within a given period of multiple days (Equation (19)) or PIN on a particular day or intraday interval (Equation (20)) are computed.

2.3 Time-Varying Probabilities of News

The estimation of PIN can be enhanced if the probabilities of news are modeled rather than assumed to be constant. Berry and Howe (1994) study the pattern of news arrival and its impact on trading volume. Using the number of news releases from Reuters' News Services as their measure of information, they report that there is "modest success" in using their information variable as an explanatory variable of trading volume on an intraday basis. This finding is concurred by Mitchell and Mulherin (1994), who use the news announcements of Dow Jones & Company. They document a positive and statistically significant relationship between public information and trading volume. More recently, Kalev et al. (2004) apply similar methodology to study the Australian market using the Signal G Database. They report that "there is statistically significant evidence that the de-trended trading volume increases as the number of news announcements per interval is higher".

The construction of an information measure based on news databases requires qualitative judgment and assessment of the proper timing of news, as discussed in great details in the above references. Furthermore, in the EHO framework, news should include both public and private information. Thus, the approach of constructing a measure of public information may not adequately proxy the probability of news. To circumvent these difficulties, we adopt a reduced-form approach. Specifically, motivated by the reported positive correlation between (public) information and trading volume, we propose to use volume as a covariate for news arrival.

We assume a logistic model in which the arrival of good news, no news, and bad news on day d depends on the aggregate *volume* of buy and sell orders. Specifically, we denote $\bar{V}^B$ as the average number of lots traded per day initiated by buy orders. Likewise, we denote $\bar{V}^S$ as the average number of lots traded per day initiated by sell orders. The numbers of lots traded on day d initiated by buy and sell orders are denoted by V_d^B and V_d^S, respectively. We then assume the probability of no news on day d to be

$$\pi_{Nd} = 1 - \theta_{Ed} = \frac{1}{1 + \exp\left\{\delta_1 + \delta_2\left[\log\left(V_d^B + V_d^S\right) - \log(\bar{V}^B + \bar{V}^S)\right]\right\}}. \tag{21}$$

We expect $\delta_2 > 0$, so that when the aggregate volume on day d, $V_d^B + V_d^S$, increases relative to the daily average volume $\bar{V}^B + \bar{V}^S$, the probability of no news decreases. Note that δ_1 is a scaling parameter, which may be positive or negative. Given news on day d, the probability of good news is assumed to be

$$\theta_{Gd} = \frac{1}{1 + \exp\left[\delta_3\left(\log V_d^S - \log \bar{V}^S\right) - \delta_4\left(\log V_d^B - \log \bar{V}^B\right)\right]}. \tag{22}$$

Again, we expect δ_3 and δ_4 to be positive, so that $V_d^S > \bar{V}^S$ or $V_d^B < \bar{V}^B$ implies a decreased probability of good news. Equation (22) implies that if $V_d^S = \bar{V}^S$ and $V_d^B = \bar{V}^B$, then the probabilities of good news and bad news, given there is news, equal one-half.

Thus, the arrivals of good news and bad news on day d are given by $\pi_{Gd} = \theta_{Ed}\theta_{Gd}$ and $\pi_{Bd} = \theta_{Ed}(1 - \theta_{Gd})$, respectively. Similar to formula (16), the likelihood function is given by

$$\prod_{d=1}^{D}\left[\sum_{s\in S}\pi_{sd}\left(\prod_{i=1}^{N_d} p_{si}(x_i, y_i|\Phi_{i-1})\right)\right], \tag{23}$$

where the probabilities π_{sd} vary with d. The MLE parameters are estimated using Equation (23). To compute PIN on day d, we use the formula

$$\text{PIN}_d = \frac{\sum_{i=1}^{N_d}\left(\pi_{Gd}\,\lambda_{1i}^{G} + \pi_{Bd}\,\lambda_{-1,i}^{B}\right)x_i}{\sum_{i=1}^{N_d}\left(\lambda_{-1,i}^{N} + \lambda_{1i}^{N} + \pi_{Gd}\,\lambda_{1i}^{G} + \pi_{Bd}\,\lambda_{-1,i}^{B}\right)x_i}, \tag{24}$$

which is a modification of Equation (20) with time-varying π_G and π_B. Again, this formula can be used to compute PIN over intraday intervals, in which case the summations in the numerator and denominator are over trades in specific intraday intervals.

3 DATA

We apply the AACD model to intraday data of five NYSE companies: Boeing (BA), General Electric (GE), International Business Machines (IBM), Altria Group (formerly Philip Morris) (MO), and AT&T (T). The data are obtained from the TAQ database for July 1, 1994 to June 30, 1995.

We extract three variables on each stock: time of trade, transaction price, and signed volume inferred using the Lee and Ready (1991) algorithm. We also correct for the opening auction and for time-of-day effects using procedures similar to those in Engle and Russell (1998). In particular, opening effects require the transactions occurring in the first 20 minutes of each day to be removed. The average duration of transactions over the following 10 minutes serves as the waiting time for the first trade after 10:00 am (EST). All transactions recorded after 4:00 pm are also deleted. In some cases, the opening transaction occurred after the first 20 minutes. Also, on a few days there are insufficient transactions between 9:50 am and 10:00 am to obtain a meaningful average starting duration. Therefore, days with opening transactions after 9:50 am and with less than three transactions over the next 10 minutes are also removed, along with November 25, 1994 due to an

Table 1 Summary statistics of duration and trade direction

Statistics	Ticker symbols				
	BA	GE	IBM	MO	T
	Average diurnally adjusted duration (in seconds)				
All trades $\overline{x}$	88.78	31.83	41.42	48.88	39.29
Buy-initiated trades $\widehat{\psi}_1$	197.86	55.23	86.31	110.29	72.29
Sell-initiated trades $\widehat{\psi}_{-1}$	161.04	75.12	79.64	87.79	86.07
	Order-flow statistics (volume in lots)				
Frequency of buys (%)	44.87	57.63	47.99	44.32	54.35
Frequency of sells (%)	55.13	42.37	52.01	55.68	45.65
Serial correlation of trade direction	0.35	0.32	0.52	0.32	0.40
Runs test of trade direction	−81.32	−132.56	−186.27	−105.77	−146.61
Average volume (lot size)	27.80	19.91	30.83	31.48	25.31
Average log volume	1.97	1.70	2.36	2.13	1.61
Average daily number of trades	243.30	677.90	521.10	442.30	549.10
Average daily number of buy-trades	109.17	390.67	250.08	196.03	298.44
Average daily number of sell-trades	134.13	287.23	271.02	246.27	250.66
Number of observations in sample	54,500	170,157	129,239	110,120	135,087

early "day after Thanksgiving" closing. Even after these deletions, a tremendous number of observations for each company remain, as documented in Table 1.

We estimate diurnal factors by applying a smoothing spline to the average duration at each time point with available data.[5] The diurnally adjusted durations are then formed by dividing each duration with the corresponding diurnal factor. For the remainder of this paper, durations x_i refer to mean-diurnally adjusted durations. The diurnal factors for all five duration series are similar to those in Engle and Russell (1998). In particular, the diurnal factors initially increase, with the largest diurnal factor occurring at the middle of the day, before decreasing.

Some summary statistics of the data are given in Table 1. The number of observations available for BA is substantially lower than the other stocks, primarily due to less frequent trading as indicated by its average duration. The average number of trades per day varies from a low of 243.3 (BA) to a high of 677.9 (GE). The runs tests indicate that there is positive serial correlation in trade directions. More than 50% of GE and T trades are buys, while the other three stocks have more sells than buys.

4 EMPIRICAL RESULTS

This section reports PIN estimates under a variety of different specifications. We also study the correlation between daily PIN estimates and daily return volatility,

[5]We use the MATLAB function *csaps.m* to compute the cubic smoothing spline. The diurnal factor is adjusted to ensure the sample mean of the diurnally-adjusted durations is equal to the sample mean of the non-diurnally-adjusted data.

Table 2 Estimates of PIN-EHO model

Variables	Parameters	Ticker symbols				
		BA	GE	IBM	MO	T
Intensity for	λ_{-1}	98.7929	345.8723	200.8416	175.2851	235.5643
sell-initiated trade		(3.2689)	(6.5457)	(6.4490)	(5.1569)	(4.8911)
Intensity for	λ_1	110.9957	269.7381	251.5821	227.1735	241.7717
buy-initiated trade		(3.6987)	(4.7169)	(6.6138)	(5.9444)	(5.6469)
Adjustment for	δ	92.6196	134.4572	148.5920	144.5634	200.0465
information		(6.8988)	(8.6544)	(8.9076)	(15.4491)	(17.5005)
Probability	θ_E	0.3511	0.4560	0.4556	0.2683	0.3539
of news		(0.0366)	(0.0426)	(0.0347)	(0.0350)	(0.0358)
Given news, probability	θ_B	0.6859	0.2870	0.2858	0.4443	0.1433
of bad news		(0.0758)	(0.0611)	(0.0684)	(0.1010)	(0.0555)
PIN-EHO		0.1342	0.0906	0.1302	0.0879	0.1292

Note. Figures in parentheses are standard errors.

as well as measures of illiquidity and market depth, to understand the economic implications of our methodology for estimating PIN.

4.1 PIN Estimates

Results of the PIN-EHO model are summarized in Table 2. This model is estimated using the MLE method with the likelihood function computed from Equation (1). PIN estimates are then computed using Equation (2). It can be seen that the PIN estimates vary from the lowest value of 0.0879 for MO to the highest value of 0.1342 for BA, which is the stock with the fewest daily trades. If we measure the relative intensity of informed traders versus uninformed traders by $2\hat{\delta}/(\hat{\lambda}_{-1} + \hat{\lambda}_1)$, the relative intensity of BA (0.883) is the highest in the sample and the relative intensity for GE (0.437) is the lowest. Thus, although BA has a lower probability of news than GE, it has a higher PIN. On the other hand, although MO has a higher relative intensity (0.718) than GE, it has the lowest PIN.

The results of the PIN–AACD model are presented in Table 3. The models are estimated using the MLE method with the likelihood function computed from Equations (15) and (16), where the conditional expected durations are defined in Equations (10) through (14). It can be seen that the estimates of the AACD models for trade direction exhibit a remarkable resemblance across the five stocks. In particular, we observe the following. First, $\hat{\nu}_{-1,-1} < \hat{\nu}_{-1,1}$ and $\hat{\nu}_{11} < \hat{\nu}_{1,-1}$ for all stocks, implying that buy trades induce lower conditional expected durations for buy trades than sell trades, and sell trades induce lower conditional expected durations for sell trades than buy trades. This property is consistent with positive serial correlation in trade direction. Second, $\hat{\varsigma}_{-1} > 0$ and $\hat{\varsigma}_1 < 0$ for all stocks, implying large buy orders induce shorter conditional expected durations for subsequent buy orders but longer conditional expected durations for sell orders. The opposite is

Table 3 Estimates of the PIN-AACD model with constant probabilities of good news, no news, and bad news

		Ticker symbols				
Trade variables	Parameters	BA	GE	IBM	MO	T
Sale after sale	$v_{-1,-1}$	1.8791	2.3126	2.9647	1.7877	1.6118
		(0.2475)	(0.0835)	(0.1096)	(0.1611)	(0.0902)
Sale after buy	$v_{-1,1}$	2.1772	2.6272	3.8651	1.9956	2.0551
		(0.2638)	(0.0885)	(0.1306)	(0.1747)	(0.1067)
Buy after sale	$v_{1,-1}$	1.7056	1.7832	2.8227	1.6843	2.1341
		(0.2711)	(0.0622)	(0.1701)	(0.1685)	(0.0890)
Buy after buy	v_{11}	1.4295	1.6645	2.1598	1.4336	1.6809
		(0.2430)	(0.0618)	(0.1432)	(0.1564)	(0.0771)
Conditional duration for sales	α_{-1}	0.5597	0.3946	0.1860	0.5079	0.5080
		(0.0445)	(0.0171)	(0.0195)	(0.0328)	(0.0207)
Lagged duration for sales	β_{-1}	0.0875	0.0722	0.1064	0.1181	0.1302
		(0.0054)	(0.0040)	(0.0067)	(0.0058)	(0.0040)
Conditional duration for buys	α_1	0.6181	0.5409	0.3776	0.5722	0.5048
		(0.0455)	(0.0137)	(0.0252)	(0.0291)	(0.0171)
Lagged duration for buys	β_1	0.1233	0.0698	0.1400	0.1473	0.1169
		(0.0059)	(0.0031)	(0.0072)	(0.0069)	(0.0040)
Adjustment for information	μ	0.5167	0.2692	0.3854	0.4351	0.4703
		(0.0275)	(0.0137)	(0.0205)	(0.0340)	(0.0300)
Probability of news	θ_E	0.3223	0.3848	0.3400	0.2070	0.2730
		(0.0388)	(0.0383)	(0.0386)	(0.0304)	(0.0425)
Given news, probability of bad news	θ_B	0.9565	0.3893	0.3754	0.7276	0.0288
		(0.0411)	(0.1213)	(0.2395)	(0.1147)	(0.0252)
Volume for sales	ς_{-1}	0.0363	0.0951	0.0201	0.0498	0.0397
		(0.0051)	(0.0036)	(0.0036)	(0.0037)	(0.0032)
Volume for buys	ς_1	–0.0466	–0.0959	–0.0421	–0.0629	–0.0373
		(0.0055)	(0.0037)	(0.0035)	(0.0044)	(0.0032)
PIN-AACD (from Equation (19))		0.1485	0.0413	0.0816	0.1011	0.1241
Daily PIN$_d$ (from Equation (20))						
Minimum		0.1336	0.0393	0.0731	0.0887	0.1082
Maximum		0.1617	0.0423	0.0890	0.1095	0.1372
Mean		0.1483	0.0411	0.0816	0.1009	0.1240

Note. Figures in parentheses are standard errors.

true for large sell orders. Thus, volume plays an explicit role in predicting trade directions.[6]

[6]To examine the fitness of the AACD models, we compute the estimated probability integral transforms (Diebold, Gunther, and Tay 1998) of the transaction durations. To conserve space, the plots are not reported here. It is observed that the distributions of the estimated probability integral transforms do not behave like a uniform distribution when the duration is very short. Otherwise, the uniform distribution

The PIN-AACD estimates are computed using Equation (19) across the entire sample period. We observe significant decreases in PIN for GE and IBM under the PIN-AACD model in comparison to the PIN-EHO model. In these two cases, estimates for the probability of news are lower under the AACD model, causing the PIN estimates to decrease. On the other hand, the PIN-AACD estimates of BA and MO increase relative to those of PIN-EHO. Indeed, for BA the μ estimate is the highest among all stocks, suggesting that the relative intensity of informed traders is high when there is news. Overall, the probability of news is similar across the five stocks, with MO and T having the least amount of news.

The bottom panel of Table 3 reports the distributions of the daily PIN-AACD estimates from Equation (20), in which the expected number of informed trades is divided by the expected number of total trades on each trading day. We report the minimum, maximum, and average daily PIN for each stock. The spreads appear to be small, indicating that there are small variations in the daily PIN under the constant probability of news assumption. Thus, while daily PIN can be estimated using transaction data, the estimates exhibit little variation when the probability of news is assumed to be constant.

Table 4 reports the MLE of the PIN-AACD models with time-varying probabilities for good and bad news, as defined in Equations (21) and (22). For brevity, only the point estimates are reported. The lower panels of Table 4 provide summary statistics for the probability of no news, good news, and bad news as well as the daily PIN estimates.

The daily PIN-AACD estimates in Table 4 are larger than their counterparts in Table 3. Moreover, they exhibit greater fluctuations over time, which is a natural outcome of allowing the probabilities of good and bad news to vary with time. Most importantly, the average probabilities of good news and bad news, across the five stocks, are more reasonable in Table 4. In particular, for each stock, the probabilities of good and bad news are similar. This symmetry implies the arrival of good versus bad news is less predictable, and consequently exerts a greater price impact upon arrival. In contrast, some of the θ_B estimates recorded in Table 3 are quite close to zero or one (see BA and T). These results highlight the advantages of a model with time-varying probabilities of news.

Figures 1 and 2 present the plots of the probabilities of good news, no news, and bad news, as well as PIN of the BA and GE data, respectively. For the BA data it can be observed that the probability of good news was quite even throughout the sample period. In contrast, the probability of bad news was higher in the latter two-thirds of the period, and in some instances was over 0.6. Consequently, PIN was found to be higher in the latter parts of the period. For the GE data, the probabilities of news appear to be quite even throughout the sample period, and the variation in PIN is smaller than that of BA.

---

appears to describe the transformations well. This finding is similar to that of Bauwens et al. (2004), who examine a wide range of error distributions of the transaction duration. We also compute the autocorrelations of the transforms up to 30 lags and find these autocorrelations to be small (less than 0.01 in absolute value).

Table 4 Estimates of the PIN-AACD model with varying probabilities of good news, no news, and bad news

		Ticker symbols				
Trade variables	Parameters	BA	GE	IBM	MO	T
Sale after sale	$v_{-1,-1}$	1.7528	2.3227	2.9234	1.6651	1.6234
Sale after buy	$v_{-1,1}$	2.0474	2.6395	3.8171	1.8652	2.0689
Buy after sale	$v_{1,-1}$	1.8132	1.7785	2.7992	1.7993	2.1314
Buy after buy	v_{11}	1.5293	1.6597	2.1413	1.5425	1.6770
Conditional duration for sales	α_{-1}	0.5794	0.3930	0.1953	0.5329	0.5059
Lagged duration for sales	β_{-1}	0.0903	0.0716	0.1074	0.1198	0.1296
Conditional duration for buys	α_1	0.6007	0.5416	0.3817	0.5505	0.5042
Lagged duration for buys	β_1	0.1221	0.0702	0.1406	0.1466	0.1175
Adjustment for information	μ	0.4934	0.2677	0.3804	0.4251	0.4686
Volume–direction for sales	ς_{-1}	0.0338	0.0948	0.0199	0.0477	0.0395
Volume–direction for buys	ς_1	−0.0483	−0.0959	−0.0420	−0.0655	−0.0372
Prob Equation, Coefficient 1	δ_1	−0.6764	−0.5004	−0.0217	−1.0131	−0.7969
Prob Equation, Coefficient 2	δ_2	2.3823	1.3632	1.9046	0.1505	0.7540
Prob equation, Coefficient 3	δ_3	1.4445	0.5441	0.0052	0.0650	−0.0067
Prob equation, Coefficient 4	δ_4	0.0234	0.5631	1.0908	0.0547	0.0016
Probability of good news						
Minimum		0.0129	0.0759	0.0148	0.1137	0.0708
Maximum		0.3655	0.3558	0.7735	0.1585	0.2930
Mean		0.1422	0.1800	0.2274	0.1315	0.1463
Std Dev		0.0698	0.0546	0.1339	0.0063	0.0378
Probability of no news						
Minimum		0.1093	0.3352	0.0806	0.6842	0.4171
Maximum		0.9855	0.8313	0.9181	0.7705	0.8577
Mean		0.6925	0.6401	0.5478	0.7369	0.7074
Std Dev		0.2053	0.1017	0.1653	0.0123	0.0751
Probability of bad news						
Minimum		0.0007	0.0764	0.0671	0.1158	0.0715
Maximum		0.7273	0.3770	0.3760	0.1573	0.2899
Mean		0.1653	0.1799	0.2249	0.1315	0.1464
Std Dev		0.1591	0.0584	0.0506	0.0064	0.0373
Daily PIN_d (from equation (24))						
Minimum		0.0109	0.0989	0.0650	0.1496	0.1021
Maximum		0.4261	0.3020	0.4258	0.1945	0.3179
Mean		0.1883	0.1889	0.2452	0.1675	0.1876
Std Dev		0.1022	0.0428	0.0676	0.0065	0.0386

We also compute PIN over intraday intervals. For each of the BA and GE stocks, we select two days for this purpose: the days with the highest and lowest probability of news. On each of these days, we consider 20-minute trade intervals and compute PIN over each interval using Equation (24), for a total of 18 PIN estimates each day. The results are presented in Figure 3. For the BA data, intraday

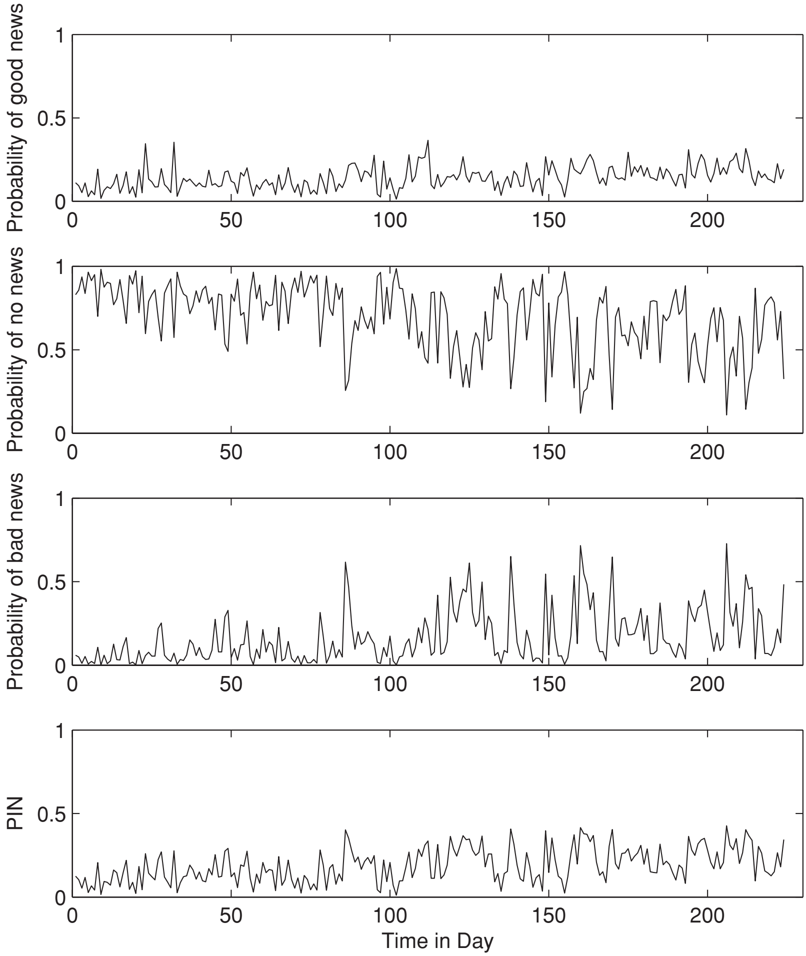

Figure 1 Probability of good news, no news, bad news, and PIN of BA data.

PIN on the day with minimum probability of news ($\hat{\theta}_E = 0.0145$) is in the interval (0.0079, 0.0142), while the interval of PIN on the day with maximum probability of news ($\hat{\theta}_E = 0.8907$) is (0.3877, 0.4662). Similarly, the intraday PIN intervals for the GE data are (0.0966, 0.1014) and (0.3014, 0.3033) for the day with minimum ($\hat{\theta}_E = 0.1687$) and maximum ($\hat{\theta}_E = 0.6648$) probability of news, respectively. Thus, intraday PIN may have substantial variations, as for the case for the BA data on a day with news (with a range of 0.0785). As the results presented here are based on a small number of stocks, it will be interesting to see if there is any regularity in intraday PIN movements over a large sample of stocks, and whether such regularities differ according to the prevalence of news.

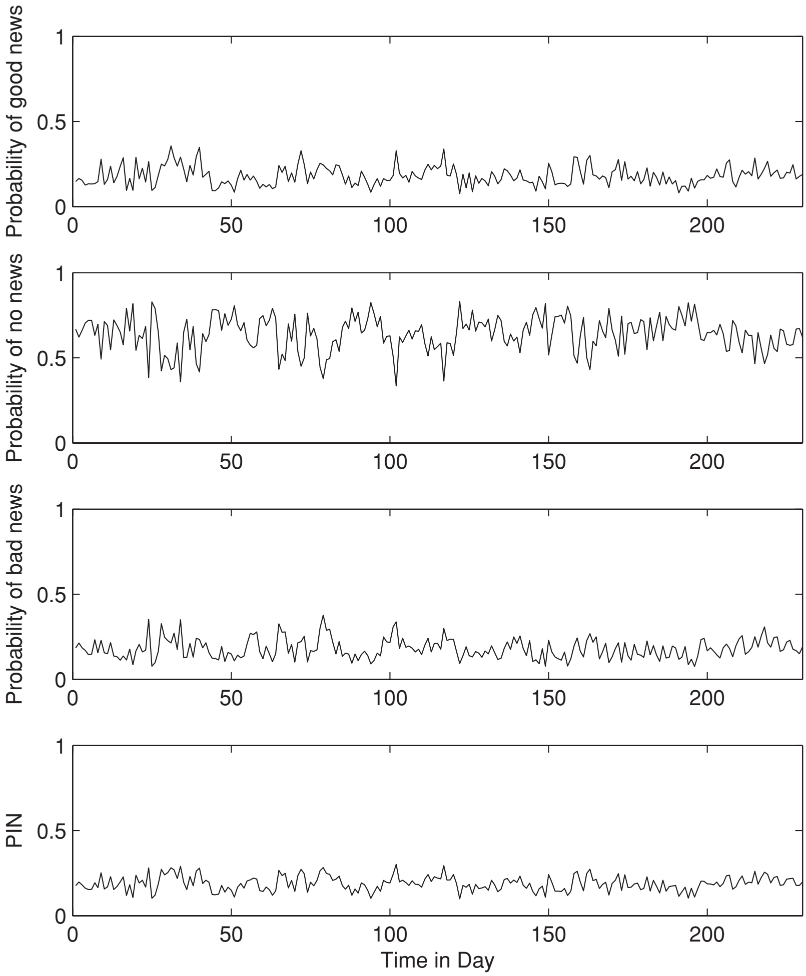

Figure 2 Probability of good news, no news, bad news, and PIN of GE data.

4.2 Evaluation of PIN Estimates

We now further examine the properties of our PIN estimates in light of the literature of the effect of asymmetric information on asset prices.

Wang (1993) demonstrates that asymmetric information increases return volatility. To this effect, if PIN successfully measures the extent of asymmetric information, we would expect PIN to correlate positively with return volatility. We denote V_d as the realized volatility, computed as the sum of squared differenced log mid-quotes. As PIN is predicted to be positively correlated with V_d, the correlation between daily PIN and daily realized volatility would help us gauge the appropriateness of our methodology, as well as the assumed specifications for the probability of no news, good news, and bad news.

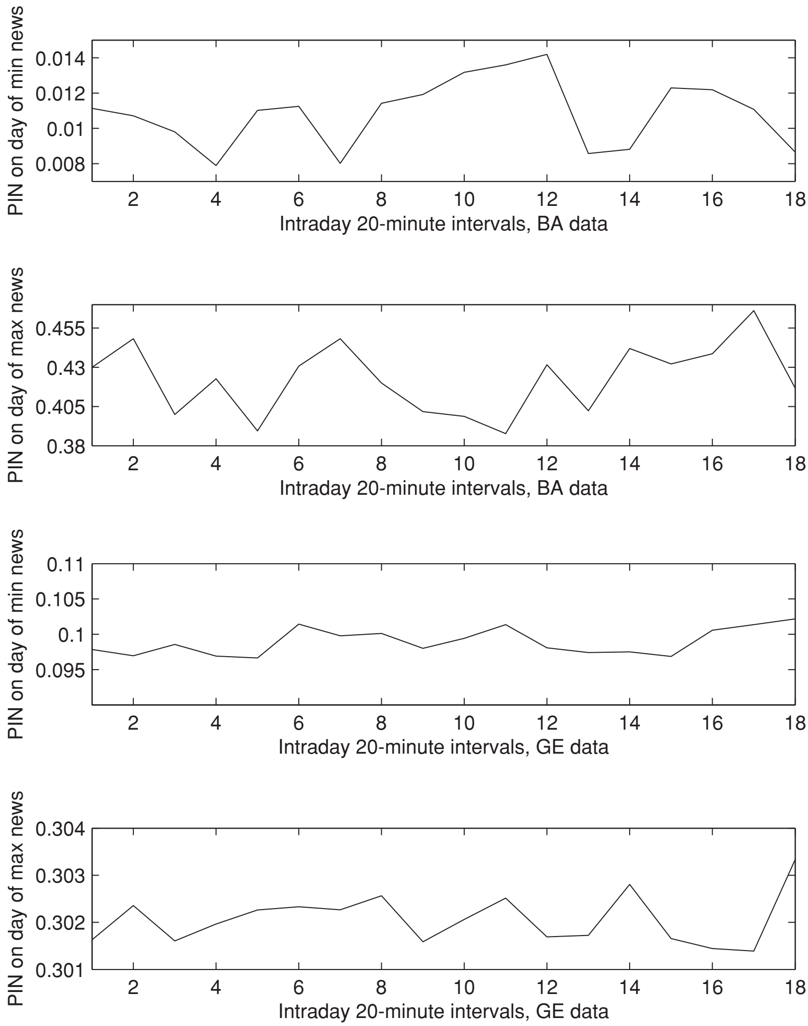

Figure 3 Intraday PIN of BA and GE data.

Another line of research investigates the effect of asymmetric information on market liquidity. Markets with higher degree of asymmetric information are expected to be less liquid and hence trade with larger spreads. Koski and Michaely (2000) focused on periods of announcements as proxy for asymmetric information and found that average spreads are higher during announcement periods. If PIN is a good measure for asymmetric information, it should be positively correlated with spread. Hence, we compute the correlation between PIN and SP, which is the average daily spread.

Anand and Weaver (2006) argued that a more comprehensive picture of liquidity is the depth of the market as measured by Kyle's (1985) lambda, which is the

Table 5 Daily correlations of PIN versus return volatility, spread, and $\hat{\lambda}_K$

Correlation	Ticker symbols				
	BA	GE	IBM	MO	T
PIN versus return volatility, with varying probability of news					
Corr(PIN_{d-1}, V_d)	0.1938	0.1746	0.1175	0.0856	0.3564
Corr(PIN_d, V_d)	**0.6001**	**0.5158**	**0.5248**	**0.4448**	**0.6722**
PIN versus return volatility, with constant probability of news					
Corr(PIN_{d-1}, V_d)	−0.0772	0.0759	0.0492	−0.0936	−0.0114
Corr(PIN_d, V_d)	0.2362	−0.0042	0.2077	0.1143	0.0517
PIN versus average spread, with varying probability of news					
Corr(PIN_{d-1}, SP_d)	0.1692	0.0581	0.1305	0.0474	0.0755
Corr(PIN_d, SP_d)	**0.3640**	**0.3443**	**0.3466**	**0.3972**	**0.2500**
PIN versus average spread, with constant probability of news					
Corr(PIN_{d-1}, SP_d)	−0.0186	−0.0651	−0.0285	0.2351	0.0004
Corr(PIN_d, SP_d)	0.1709	−0.1746	0.1850	0.3846	−0.0266
PIN versus Kyle's lambda, with varying probability of news					
Corr(PIN_{d-1}, $\hat{\lambda}_{Kd}$)	0.0698	0.0689	0.1021	0.0886	0.0639
Corr(PIN_d, $\hat{\lambda}_{Kd}$)	**0.1546**	**0.1593**	**0.2164**	**0.3689**	**0.0807**
PIN versus Kyle's lambda, with constant probability of news					
Corr(PIN_{d-1}, $\hat{\lambda}_{Kd}$)	−0.0022	−0.0032	−0.0180	0.1276	−0.1569
Corr(PIN_d, $\hat{\lambda}_{Kd}$)	0.0533	−0.0691	0.1830	0.2673	−0.0346

Note. V_d is the integrated volatility, SP_d is the average spread, and $\hat{\lambda}_{Kd}$ is Kyle's lambda.

price movement per unit signed volume of trade. We follow Anand and Weaver's (2006) method and estimate Kyle's lambda $\hat{\lambda}_K$ as the regression coefficient of price changes on signed logarithmic volumes. Large values of $\hat{\lambda}_K$ reflect the lack of depth of the market. Our regressions are based on transaction data over 15-minute intervals and $\hat{\lambda}_K$ are computed daily. We then calculate the correlation coefficients between daily $\hat{\lambda}_K$ and estimates of PIN. If PIN reflects asymmetric information, there should be positive correlation between PIN and $\hat{\lambda}_K$.

Table 5 shows that time-varying probabilities of news yield much higher correlations between V_d and the contemporaneous (as well as lagged) PIN. Indeed, the contemporaneous correlations, Corr(PIN_d, V_d), range from 0.4448 to 0.6722. In contrast, under the assumption that the probabilities of good news and bad news are constant, the contemporaneous correlation for GE is negative, while the maximum correlation among other stocks is 0.2362 (for BA). We also observe that the contemporaneous correlations between PIN and average spread, Corr(PIN_d, SP_d), are all positive for the model with time-varying probabilities. When the model with constant probability of news is used, two of the stocks have negative values of Corr(PIN_d, SP_d). Finally, the contemporaneous correlations of PIN with $\hat{\lambda}_{Kd}$ are all positive when PIN is computed from the model with time-varying probabilities of news. When constant probabilities of news are assumed, the correlations are lower and may be negative. To sum up, daily PIN based on time-varying probabilities of news significantly correlates with contemporaneous volatility, spread, and Kyle's lambda at the 5% level of significance (except for Stock T for the correlation with Kyle's lambda; see the figures in boldface in Table 5).

Overall, the results in Table 5 support our proposed extension of the EHO formulation that allows for time-varying probabilities of good news and bad news.

5 CONCLUSIONS

This paper extends the AACD model to incorporate trade directions, volume, durations, and their interactions into modeling the transaction data of buy- and sell-initiated orders. The model is then used to estimate the probability of informed trading. The use of transaction data allows us to relax the assumption of independent aggregate buy and sell orders in the EHO framework.

Our enhanced methodology yields daily estimates for the probability of informed trading, and allows the underlying probabilities of good news and bad news to be time varying. A comparison of our daily estimates for the probability of informed trading with return volatility verifies the improved economic implications of our methodology.

Future research may utilize our daily estimates of the probability of informed trading to study the impact of various events such as earnings announcements or merger activity on the level of informed trading. More generally, the asset pricing implications of informed trading may be investigated with greater precision using our methodology.

APPENDIX

We consider a two-parameter Weibull distribution for the inter-arrival times of the latent processes with the following probability density function:

$$f_T(t) = \frac{\phi}{\psi}\left(\frac{t}{\psi}\right)^{\phi-1} \exp\left[-\left(\frac{t}{\psi}\right)^{\phi}\right],$$

where ψ and ϕ are, respectively, the scale and shape parameters, with $\psi > 0$ and $\phi > 0$. Denoting the ψ_{ji} as the conditional scale parameter of trade direction j given information Φ_{i-1} and ϕ_j as the shape parameter for the latent trade direction $j = -1, 1$, the conditional joint density function $p_i(x_i, y_i|\Phi_{i-1})$ of x_i and y_i is then given by

$$p_i(x_i, k|\Phi_{i-1}) = \prod_{j=-1,1}\left[\frac{\phi_j}{\psi_{ji}}\left(\frac{x_i}{\psi_{ji}}\right)^{\phi_j-1}\right]^{D_k(j)} \exp\left[-\left(\frac{x_i}{\psi_{ji}}\right)^{\phi_j}\right], \qquad k = -1, 1.$$

Given a sample of observations $\{x_i, y_i\}$ for $i = 1, \ldots, N$, the log-likelihood function is

$$\sum_{i=1}^{N} \log\ p_i(x_i, y_i | \Phi_{i-1}) = -\sum_{i=1}^{N} \left\{ \sum_{j=-1,1} \left(\frac{x_i}{\psi_{ji}} \right)^{\phi_j} - \sum_{j=-1,1} D_{y_i}(j) \log \left[\frac{\phi_j}{\psi_{ji}} \left(\frac{x_i}{\psi_{ji}} \right)^{\phi_j - 1} \right] \right\}.$$

Thus, the parameters of the model can be estimated using the MLE method once the functional forms of the scale parameters ψ_{ji} are specified.

If the shape parameters of the two latent competing processes of trade directions are equal, i.e., $\phi_{-1} = \phi_1 = \phi$, say, the conditional joint density function of x_i and y_i is then given by

$$p_i(x_i, k | \Phi_{i-1}) = \frac{\phi}{\psi_{ki}} \left(\frac{x_i}{\psi_{ki}} \right)^{\phi - 1} \exp \left[-\sum_{j=-1,1} \left(\frac{x_i}{\psi_{ji}} \right)^{\phi} \right], \qquad k = -1, 1.$$

In this special case it is straightforward to obtain the marginal densities of x_i and y_i. Thus, if we define ψ_i by

$$\psi_i = \left(\frac{1}{\psi_{-1,i}^{\phi}} + \frac{1}{\psi_{1i}^{\phi}} \right)^{-\frac{1}{\phi}},$$

the conditional marginal density of x_i is

$$f_{x_i}(x | \Phi_{i-1}) = \frac{\phi x^{\phi - 1}}{\psi_i^{\phi}} \exp \left[-\left(\frac{x}{\psi_i} \right)^{\phi} \right].$$

Also, the conditional marginal density of y_i is

$$f_{y_i}(k | \Phi_{i-1}) = \left(\frac{\psi_i}{\psi_{ki}} \right)^{\phi}, \qquad k = -1, 1.$$

Hence, conditional on Φ_{i-1}, x_i has a two-parameter Weibull distribution with shape parameter ϕ and scale parameter ψ_i. Likewise, conditional on Φ_{i-1}, y_i has a multinomial distribution with probabilities proportional to $1/\psi_{ki}$ for $k = -1, 1$.

Note that the joint density of x_i and y_i is equal to the product of the densities of x_i and y_i. Thus, under the special case when the shape parameters of the latent processes are equal, the trade direction y_i and the trade duration x_i are independent conditional upon the information Φ_{i-1}.

Given the shape parameter ϕ and the scale parameter ψ_{ki} for $k = -1, 1$, the expected duration of the next arrival is

$$\psi_{ki}\, \Gamma \left(1 + \frac{1}{\phi} \right),$$

where $\Gamma(\cdot)$ is the gamma function. We take the reciprocal of the above expression as the conditional expected number of trade per unit time. Thus, using Equation (3) to compute PIN, the terms involving ϕ drop out and we may use Equation (24) to compute PIN, with λ_{ki}^{s} replaced by the reciprocal of ψ_{ki}^{s}, for $k = -1, 1$ and $s \in \{G, N, B\}$.

Received February 12, 2008; revised March 4, 2009; accepted March 20, 2009.

REFERENCES

Aktas, N., E. de Bodt, F. Declerck, and H. Van Oppens. 2007. "The PIN Anomaly around M&A Announcements." *Journal of Financial Markets* 10: 169–191.

Anand, A., and D. G. Weaver. 2006. "The Value of the Specialist: Empirical Evidence from the CBOE." *Journal of Financial Markets* 9: 100–118.

Aslan, H., D. Easley, S. Hvidkjaer, and M. O'Hara. 2006. "The Determinants of Informed Trading: Implications for Asset Pricing" Working Paper, Cornell University.

Bauwens, L., and P. Giot. 2000. "The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks." *Annales d'Économie et de Statistique* 60: 117–149.

Bauwens, L., and P. Giot. 2003. "Asymmetric ACD Models: Introducing Price Information in ACD Models." *Empirical Economics* 28: 709–731.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas. 2004. "A Comparison of Financial Duration Models via Density Forecasts." *International Journal of Forecasting* 20: 589–609.

Bauwens, L., and N. Hautsch. 2006. "Stochastic Conditional Intensity Process." *Journal of Financial Econometrics* 4: 450–493.

Bauwens, L., and D. Veredas. 2004. "The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Durations." *Journal of Econometrics* 119: 381–412.

Benos, E., and M. Jochec. 2007. "Testing the PIN Variable." Working Paper, University of Illinois.

Berry, T. D., and K. M. Howe. 1994. "Public Information Arrival." *Journal of Finance* 49: 1331–1346.

Bisière, C., and T. Kamionka. 2000. "Timing of Orders, Order Aggressiveness and the Order Book at the Paris Bourse." *Annales d'Économie et de Statistique* 60: 43–72.

Chung, K. H., and M. Li. 2003. "Adverse-Selection Costs and the Probability of Information-Based Trading." *Financial Review* 38: 257–272.

Diamond, D. W., and R. Verrecchia. 1991. "Disclosure, Liquidity, and the Cost of Capital." *Journal of Finance* 66: 1325–1359.

Diebold, F. X., T. A. Gunther, and A. S. Tay. 1998. "Evaluating Density Forecasts with Applications to Financial Risk Management." *International Economic Review* 39: 863–883.

Duarte, J., and L. Young. 2009. "Why Is PIN Priced?" *Journal of Financial Economics* 91: 119–138.

Easley, D., S. Hvidkjaer, and M. O'Hara. 2002. "Is Information Risk a Determinant of Asset Returns?" *Journal of Finance* 57: 2185–2221.

Easley, D., R. F. Engle, M. O'Hara, and L. Wu. 2008. "Time-Varying Arrival Rates of Informed and Uninformed Trades." *Journal of Financial Econometrics* 6: 171–207.

Easley, D., N. Kiefer, and M. O'Hara. 1996. "Cream-Skimming or Profit Sharing? The Curious Role of Purchased Order Flow." *Journal of Finance* 51: 811–833.

Easley, D., N. Kiefer, and M. O'Hara. 1997. "The Information Content of the Trading Process." *Journal of Empirical Finance* 4: 159–186.

Easley, D., N. Kiefer, M. O'Hara, and J. B. Paperman. 1996. "Liquidity, Information, and Infrequently Traded Stocks." *Journal of Finance* 51: 1405–1436.

Easley, D., M. O'Hara, and G. Saar. 2001. "How Stocks Splits Affect Trading: A Microstructure Approach." *Journal of Financial and Quantitative Analysis* 36: 25–51.

Engle, R. 2000. "The Econometrics of Ultra-High Frequency Data." *Econometrica* 68: 1–22.

Engle R., and J. R. Russell. 1998. "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data." *Econometrica* 66: 1127–1162.

Fernandes, M., and J. Grammig. 2005. Non-parametric Specification Tests for Conditional Duration Models. *Journal of Econometrics* 127: 35–68.

Ghysels, E., C. Gourieroux, and J. Jasiak. 2004. "Stochastic Volatility Duration Models." *Journal of Econometrics* 119: 413–435.

Grammig, J., and K. O. Maurer. 2000. "Non-monotonic Hazard Functions and the Autoregressive Conditional Duration Model. *Econometrics Journal* 3: 16–38.

Hautsch, N. 2004. "Modelling Irregularly Spaced Financial Data." Berlin: Springer.

Henry, T. R. 2006. "Short Selling, Informed Trading, and Stock Returns." Working Paper, University of Georgia.

Kalev, P. S., W-M. Liu, P. K. Pham, and E. Jarnecic. 2004. "Public Information Arrival and Volatility of Intraday Stock Returns." *Journal of Banking and Finance* 28: 1441–1467.

Koski, J. L., and R. Michaely. 2000. "Prices, Liquidity, and the Information Content of Trades." *Review of Financial Studies* 13: 659–696.

Kyle, A. 1985. "Continuous Auctions and Insider Trading." *Econometrica* 53: 1315–1335.

Lee, C. M. C., and M. A. Ready. 1991. "Inferring Trade Direction from Intraday Data." *Journal of Finance* 46: 733–746.

Lei, Q., and G. Wu. 2005. "Time-Varying Informed and Uninformed Trading Activities." *Journal of Financial Markets* 8: 153–181.

Lo, I., and S. G. Sapp. 2005. "Order Submission: The Choice between Limit and Market Orders." Bank of Canada Working Paper 2005-42.

Mitchell, M. L., and J. H. Mulherin. 1994. "The Impact of Public Information on the Stock Market." *Journal of Finance* 49: 923–950.

Pacurar, M. 2008. "Autoregressive Conditional Duration (ACD) Models in Finance: A Survey of the Theoretical and Empirical Literature." *Journal of Economic Surveys* 22: 711–751.

Russell, J. R. 1999. "Econometric Modeling of Multivariate Irregularly-Spaced High-Frequency Data." Working Paper, University of Chicago.

Wang, J. 1993. "A Model of Intertemporal Asset Prices under Asymmetric Information." *Review of Economic Studies* 60: 249–282.

Zhang, M. Y., J. R. Russell, and R. S. Tsay. 2001. "A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data. *Journal of Econometrics* 104: 179–207.